DOCUMENT RESUME

ED 442 827 TM 031 247

AUTHOR Shen, Linjun
TITLE Curriculum Type as a Differentiating Factor in Medical Licensing Examinations.
PUB DATE 2000-04-00
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Curriculum; *Item Bias; *Licensing Examinations (Professions); *Medical Students; *Osteopathy; Test Items

ABSTRACT

This study assessed the effects of the type of medical curriculum on differential item functioning (DIF) and group differences at the test level in Level 1 of the Comprehensive Osteopathic Medical Licensing Examinations (COMLEX). The study also explored the relationship of the DIF and group differences at the test level. There are generally two types of osteopathic preclinical curriculum: traditional discipline-based and organ system-based. The June 1998 administration of COMLEX was used. This examination involved 800 multiple-choice items and 2,122 students from 17 osteopathic schools. Of these students, 578 studied an organ-system-based curriculum. The study reveals that the directions of the influences of curriculum type were consistent at the test and item levels, but the magnitude of the influences was minimal. At the total examination level, students from the system-based curriculum performed better but with no practical significance. At the item level, the majority of items were immune from the impact of curriculum, while more items of those that had significant DIF favored students in organ-system based curricula. It is concluded that the advantages of students in the system-based curricula on the total examination were due in part to items discriminative against students in the discipline-based curricula. (Contains 2 tables and 11 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 827

L. Shen

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Curriculum Type as a Differentiating Factor in Medical Licensing Examinations

Linjun Shen, Ph.D., MPH

TM031247

A paper presented at the annual conference of American Educational Research Association, April 2000, New Orleans

2

BEST COPY AVAILABLE

Curriculum Type as a Differentiating Factor in Medical Licensing Examinations

Research on the differentiating factors in medical licensure examinations has focused on, among others, gender and ethnic groups[1,2], content appropriateness of certain disciplines or topics[3,4,5,6], and types of curriculum[7]. Methodologically, the majority of these analyses concentrated on the differentiating effects at the whole test or subtest level. Few studies addressed the effects of differentiating factors at the item level in medical licensing examinations[8]. Group differences at the test level and the item level are different phenomena and involve different issues. It is not clear if group differences at the test level would guarantee the same differences at the item level, or if an individual item's differentiating function would cause the group differences at the test level. To maintain a high standard in medical licensure examinations, it is necessary to address the issues of differential item functioning (DIF) in medical licensing examinations and the relationship between DIF and group differences at the test level . This paper assessed the effects of the type of medical curricula on DIF and the group differences at the test level in Level 1 of Comprehensive Osteopathic Medical Licensing Examinations (COMLEX) and explored the relationship of the DIF and the group differences at the test level.

Traditionally, Part I of osteopathic board examination was constructed based on disciplines. The newly-developed COMLEX Level 1 broke the boundaries among disciplines and constructed the exam around clinical topics[9]. Therefore, success on COMLEX Level 1 requires integration of disciplines and application of basic biomedical science knowledge into clinical scenarios. There are generally two types of osteopathic preclinical curriculum: traditional discipline-based curriculum and organ system-based curriculum. Intuitively, organ-system-based curriculum would be more advantageous on the COMLEX Level 1 over traditional curriculum since it is inter-disciplined and more clinical. Curriculum is an educational factor and will inevitably influence students' achievement. It is, therefore, not surprising if students of the two types of curricula perform differently on the Level 1 as a whole test. However, there has never been evidence that one type of curriculum is systematically superior to another. Therefore, it is not desirable if all or the majority of the items in the medical licensing examinations favor one type of curriculum over another. From this perspective, this paper also investigated the pattern and the scope of DIF due to curriculum type in COMLEX Level 1.

Methods

Instrumentation and subjects

The examination involved in this study was the June 1998 administration of COMLEX Level 1. The COMLEX is a three-level series of licensing examinations for osteopathic physicians. Level 1 is constructed according to two dimensions: clinical presentations and physician tasks, emphasizing scientific understanding of health or disease mechanisms. Level 1 covers seven preclinical disciplines: anatomy, physiology, biochemistry, pharmacology, pathology, microbiology, and osteopathic principles.

The reliability of the June 1998 Level 1 exam was .97. That exam had 800 multiple-choice items and 2356 students from 17 osteopathic schools, of which 2120 were first-time takers. This

BEST COPY AVAILABLE

analysis only used first-time takers as subjects. Based on characteristics of preclinical curriculum, 11 schools with 1542 first-timers were considered as discipline-based curriculum schools, the other 6 with 578 first-timers were organ-system-based.

A total of 70 items, 10 for each of the disciplines, were randomly selected from the exam for the DIF analysis. The selected items covered a broad range of clinical topics.

Analysis

The group difference on the total exam was estimated by a multiple linear regression with the Rasch-model-based person measure logit on the whole exam as the dependent variable and a dichotomous variable of curriculum type as the independent variable. In addition, a dichotomous variable of sector was fitted into the model as a covariate. A previous study reported that sector was a significant predictor of osteopathic student performance on the osteopathic licensing exams[10]. Therefore, the effects of curriculum type in this study were adjusted for the effects of sector.

Students' responses to individual items were functions of their overall achievement level. Theoretically, more capable students were more likely to get any particular item correct. So, observed response difference on an item between two groups might not the pure effects of curriculum type and were likely confounded by the overall achievement level. To be consistent with the analysis of group differences on the total exam, effects of sector also need to be taken into account in assessing curriculum effects. For these considerations, the criterion of DIF in this study was that if the probability of getting an item correct for one group of students was significantly different from that of another group after controlling the level of overall achievement and effects of sector, the item would be considered as having DIF. Two approaches were used to achieve this consideration: Mantel-Haenszel (MH) method and logistic regression. Using two methods for the same data allowed comparisons of both the results and effectiveness of the two methods.

With the Mantel-Haenszel method (MH), subjects were divided into 7 approximately equally-sized groups based on the Rasch-model person measures on the whole exam. At each achievement level, subjects were further divided into two groups according to public or private schools. At each of the 14 achievement level-by-sector strata, a two-by-two table was constructed to represent the association between two types of responses to an item (correct or incorrect) and two types of curricula. Correct response was coded as 1 and incorrect as 0, while organ-system curriculum was coded as 1 and discipline-based as 0. This stratification allowed the association between item response type and curriculum type to be adjusted for the overall achievement level and sector. The MH method generated a point estimate of an overall odds ratio, a pooled summary estimate from the 14 two-by-two tables, as well as a 95% confidence interval of the ratio, to describe the magnitude of the association between response type and curriculum type adjusted for the achievement level and sector. According to the coding of this study, an odds ratio with a low bound of confidence interval greater than 1.00 would indicate that the curriculum type was significantly associated with students' response type and students in organ-system based curriculum were more likely to get an item correct. An odds ratio with the

up-bound of confidence interval smaller than 1.00 would indicate a significant inverse association with students in a discipline-based curricula having a greater advantage. The MH method also calculated a chi-square statistic to test the significance of the overall association by pooling information of the 14 strata.

Using the logistic regression approach, this analysis fitted a linear regression model, with the log odds of success on an item being the dependent variable. The independent variables were Rasch-model-based personal measure on the whole exam, a variable of sector, and the dichotomous variable of curriculum type. As with the MH method, the effect of curriculum type estimated by the logistic regression was adjusted for the overall achievement level and sector. Similarly, the parameter estimated for the curriculum type was tested for significance and was converted into an odds ratio with a 95% confidence interval to represent the relative likelihood for students in an organ-based curriculum getting an item correct. Following the ETS standards of DIF as specified by Holland[11], this study adopted .01 as the significance level of the DIF test.

Results

The observed mean logits on the total exam for 578 students in an organ-system-based curriculum and 1542 students in a discipline-based curriculum were .988 and .905 respectively. The regression coefficient for the variable of curriculum type after adjusting for sector was .0844, significantly different from zero with a p-value smaller than .0001. This indicated that, overall, students educated by an organ-system based curriculum performed better on COMLEX Level 1 than their peers in traditional curriculum. Nevertheless, the R square of the model was only .01, indicating the curriculum type and sector of schools together could only explain 1% of the variations of the COMLEX test results among candidates. Thus, even statistically students with organ-based curriculum did better than those with discipline-based curriculum, practically the effects of curriculum type was minimal and curriculum type was not decisive factor for candidates' performance on the COMLEX Level 1.

The DIF analysis results from the MH method and logistic regression were almost identical. Table 1 details the findings by the two methods on all 70 items. Table 2 summarizes the results by disciplines. As Table 2 indicates, both methods discovered that 17 items out of the 70 randomly selected items had significant DIF. Among the 17 items with significant DIF, 11 items favored organ-system curriculum students while the other 6 favored students in traditional curriculum.

A review of psychometric properties of the 17 items suggested that quality of those items was acceptable. No item had a discrimination index smaller than .15 or p-value smaller than .35. The 17 items were also sent to two content experts to review. The reviewers were asked to focus on two issues: (1) are contents of the items reasonable and necessary to have been included in the COMLEX Level 1? (2) are there any flaws, such as topic, wording, etc., in the items which would discriminate candidates in either type of curricula? Except one reviewer thought the topic of one anatomy item was not necessarily crucial, both reviewers confirmed that contents of those items were important for medical practice. Further they did not find content or technical flaws which would obviously benefit either of the curricula.

The size of the odds ratios of the 17 items was remarkably small. For 11 items favoring organ-system curriculum, only two items had odds ratios barely greater than 2.00. For the rest of the 6 items, three had an up-bound of the confidence interval greater than .90.

Discussion

This study revealed that the directions of the influences of curriculum type were consistent at the test and item levels, but the magnitude of the influences was minimal. At the total exam level, students in system-based curricula performed better but with no practical significance. At the item level, the majority of items were immune from the impact of curriculum, while more items of those with significant DIF favored students in organ-system based curricula. Therefore, it is save to conclude that the advantages of the students in the system-based curricula on the total exam of the COMLEX Level 1 was partly due to items discriminative against students in the discipline-based curricula.

It is not surprising to find that the organ-system based curricula did better on COMLEX Level 1 and, among items with DIF, more items favored that type of curricula. It is also encouraging to find that the size of the curriculum influences was controlled at the minimum level. DIF is not a desirable item property for any medical licensing examinations. However, DIF caused by educational factors is not necessarily equivalent to content bias especially for items with a marginal significance level. The ultimate judgment of the validity of DIF items is an item-by-item review by content experts. If the content is not biased, the total number of DIF items is relatively small, and there is no clear single direction among the DIF items, DIF items due to some educational factors can be kept in the exam. A comprehensive review of the items with DIF needs to consider effects of disciplines, specific clinical topics, and psychometric features of those items. Moreover, each item has its own specific topic, therefore, has its unique determinants. Interpretation of DIF has to be item specific. At the same time, the same items may have DIF due to other factors. So, the more factors to be controlled, the better assessment of DIF there will be. On the other hand, efforts are still needed to reduce the number of DIF items. Further studies involving careful comparisons of items with and without DIF should help in constructing items with less chance of having DIF.

References

1. Dawson B, et al. Performance on the National Board of Medical Examiners Part I Examination by men and women of different race and ethnicity. JAMA, 1994, 272:674-9.

2. Bestler, J.M. Ethnicity, sex, and the NBME Part I examination. JAMA, 1995,273:702.

3. Ruckdeschel, JC, et al. Content bias in the neoplastic-related items of the National Board of Medical Examiners Part II Examination. Medical Pediatric Oncology, 1982,10:269-73.

4. Sanders, AB, et al. An analysis of Part II of the National Board of Medical Examiners Test with regard to questions in emergency medicine. Annuli of Emergency Medicine, 1986, 15:12-5.

5. Hark LA, et al. Nutrition coverage on medical licensing examinations in the United States. American Journal of Clinical Nutrition, 1997, 65:568-71.

6. Murrage SF, et al. Comparing the transfusion medicine content of the NBME's examinations in 1984-1985 and 1989-1990. Academic Medicine, 1992,67:608-10.

7. Mennin S.P., et al. Performances on the NBME I, II, and III by medical students in the problem-based learning and conventional tracks at the University of New Mexico. Academic Medicine, 1993,68:616-24.

8. Sykes, Robert C, et al. Dimensionality and DIF in a licensure examination. A paper presented at the annual conference of American Educational Research Association, April 1993, Chicago.

9. Bulletin of Information, National Board of Osteopathic Medical Examiners, 1999.

10. Shen, L. The growth pattern of general medical knowledge. In Proceedings of the Six Ottawa Conference on Medical Education, 1994

11.Holland, P. W., Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. Braun (Ed.), Test Validity, Hillsdale, NJ: Lawrence Erlbaum Associates.

Table 1

Results of DIF Analysis

Item	P value MH[a]	P value LogR[b]	Odds Ratio MH	Odds Ratio LogR	95% C.I. of Odds Ratio MH	95% C.I. of Odds Ratio LogR
			Anatomy			
1	.434	.386	1.083	1.093	.887, 1.323	0.894, 1.337
2	.498	.456	1.256	1.292	.650, 2.429	0.659, 2.534
3	.083	.074	1.343	1.363	.963, 1.875	0.971, 1.914
4	.045	.032	1.220	1.237	1.004, 1.483	1.019, 1.502
5	.383	.318	1.086	1.099	.903, 1.306	0.913, 1.321
6	.002	.004	0.752	0.763	.626, 0.904	0.633, 0.919
7	.261	.304	0.829	0.842	.598, 1.149	0.606, 1.169
8	.823	.042	0.823	0.822	.682, .993	0.681, 0.993
9	.086	.099	1.178	1.170	.977, 1.420	0.971, 1.410
10	.013	.016	0.782	0.787	.643, 0.950	0.648, 0.955
			Physiology			
1	.007	.011	1.317	1.299	1.077, 1.611	1.063, 1.588
2	.006	.010	0.761	0.775	0.627, 0.923	0.639, 0.940
3	.397	.402	1.409	1.413	0.637, 3.117	0.630, 3.171
4	.003	.002	0.748	0.741	0.620, 0.904	0.613, 0.897
5	.010	.014	0.653	0.664	0.472, 0.904	0.479, 0.920
6	.001	.0001	0.463	0.470	0.382, 0.560	0.387, 0.572
7	.871	.894	1.016	1.013	0.835, 1.238	0.833, 1.233
8	.001	.0001	0.421	0.420	0.339, 0.522	0.337, 0.523
9	.948	.933	0.988	0.985	0.697, 1.402	0.692, 1.402
10	.027	.039	0.780	0.793	0.626, .973	0.636, 0.989
			Biochemistry			
1	.555	.548	0.767	0.764	0.318, 1.849	0.317, 1.841
2	.001	.0001	0.635	0.642	0.525, 0.767	0.530, 0.776
3	.102	.075	1.184	1.201	0.967, 1.449	0.982, 1.469
4	.023	.029	.782	0.782	0.632, 0.967	0.628, 0.975
5	.001	.0001	1.637	1.642	1.352, 1.984	1.355, 1.990
6	.001	.0001	1.711	1.714	1.321, 2.217	1.322, 2.221
7	.002	.002	1.451	1.470	1.141, 1.844	1.155, 1.872

8	.351	.373	1.096	1.091	0.905, 1.327	0.901, 1.320
9	.928	.961	0.991	0.995	0.823, 1.195	0.827, 1.198
10	.184	.195	1.138	1.133	0.940, 1.376	0.938, 1.370

Pharmacology

1	.004	.007	0.673	0.685	0.515, 0.881	0.522, 0.899
2	.032	.038	0.802	0.807	0.655, 0.982	0.659, 0.988
3	.984	.974	1.002	0.997	0.826, 1.216	0.821, 1.210
4	.553	.427	1.071	1.091	0.864, 1.327	0.881, 1.351
5	.946	.772	1.009	1.041	0.773, 1.317	0.795, 1.362
6	.324	.389	0.907	0.918	0.746, 1.102	0.755, 1.116
7	.855	.726	1.027	1.052	0.775, 1.360	0.794, 1.394
8	.042	.058	0.688	0.707	0.480, 0.987	0.494, 1.012
9	.020	.022	0.801	0.803	0.664, 0.966	0.665, 0.969
10	.110	.147	0.615	0.643	0.339, 1.115	0.354, 1.168

Pathology

1	.001	.0001	1.75	1.779	1.385, 2.213	1.405, 2.251
2	.423	.382	0.925	0.918	0.765, 1.119	0.759, 1.111
3	.137	.105	1.366	1.413	0.905, 2.062	0.930, 2.146
4	.001	.0009	1.386	1.376	1.146, 1.676	1.139, 1.663
5	.364	.459	0.917	0.932	0.761, 1.106	0.774, 1.123
6	.557	.561	0.920	0.921	0.697, 1.215	0.697, 1.216
7	.110	.138	1.167	1.152	0.965, 1.411	0.956, 1.389
8	.001	.0001	2.116	2.180	1.551, 2.887	1.590, 2.988
9	.463	.455	0.933	0.932	0.774, 1.123	0.774, 1.121
10	.001	.0001	2.081	2.120	1.507, 2.875	1.526, 2.946

Microbiology

1	.001	.0003	1.414	1.454	1.155, 1.730	1.187, 1.780
2	.001	.0003	1.446	1.442	1.192, 1.755	1.185, 1.754
3	.626	.643	1.049	1.046	0.866, 1.270	0.865, 1.265
4	.273	.347	0.828	0.847	0.591, 1.161	0.599, 1.197
5	.756	.756	1.032	1.032	0.847, 1.256	0.847, 1.257
6	.135	.150	0.866	0.871	0.716, 1.046	0.721, 1.051
7	.589	.605	0.951	0.953	0.794, 1.140	0.795, 1.143
8	.118	.138	0.839	0.847	0.674, 1.045	0.680, 1.055
9	.001	.0001	1.652	1.656	1.331, 2.050	1.374, 2.054
10	.828	.888	0.972	0.982	0.755, 1.253	0.761, 1.267

(Table 1 continued)

OPP 1

1	.725	.727	1.047	1.047	0.810, 1.354	0.808, 1.359
2	.040	.044	.823	0.825	0.683, 0.991	0.685, 0.995
3	.417	.446	1.106	1.101	0.866, 1.413	0.860, 1.410
4	.004	.003	1.314	1.327	1.093, 1.581	1.100, 1.602
5	.118	.175	0.860	0.818	0.712, 1.039	0.727, 1.060
6	.173	.136	1.165	1.181	0.935, 1.450	0.949, 1.470
7	.098	.079	1.200	1.213	0.967, 1.490	0.978, 1.505
8	.143	.122	1.167	1.178	0.949, 1.436	0.957, 1.451
9	.582	.648	0.894	0.911	0.599, 1.334	0.611, 1.359
10	.772	.771	1.034	1.034	0.825, 1.296	0.825, 1.296

a. Mantel-Haenszel approach

b. Logistic regression approach

Table 2

Curriculum Type and Items with Differential Item Function

		Number of Items with DIF	
Discipline	Total # with DIF	Favoring Organ-System	Favoring Traditional
Anatomy	1	0	1
Physiology	3	0	3
Biochemistry	4	3	1
Pharmacology	1	0	1
Pathology	4	4	0
Microbiology	3	3	0
Osteopathic Principles	1	1	0
Total	17	11	6